Survival Analysis / Reliability Analysis / Churn Modeling

X, a rv, is a survival model if S_X ≥ 0 and no maximum.

Some brand name discrete survival models are:
geometric, extnegbinomial, poisson

Some brand name continuous survival models are:
gamma, lognormal, weibull**, pareto, F, betaprime $\quad k, \lambda > 0$

Review the Weibull. Let $Y_1, \ldots, Y_n \overset{iid}{\sim} \text{Weibull}(k, \lambda) := \lambda k (\lambda y)^{k-1} e^{-(\lambda y)^k} \mathbb{1}_{y \geq 0}$

$$F(y) = 1 - e^{-(\lambda y)^k} \Longleftrightarrow S(y) := 1 - F(y) = P(Y > y) = e^{-(\lambda y)^k}$$

$$\theta := E[Y] = \int_0^\infty y \left(\lambda k (\lambda y)^{k-1} e^{-(\lambda y)^k}\right) dy = \lambda^k k \int_0^\infty y^k e^{-\lambda^k y^k} dy$$

let $u = \lambda^k y^k \Rightarrow y^k = \frac{u}{\lambda^k} \Rightarrow y = \frac{u^{\frac{1}{k}}}{\lambda} \Rightarrow \frac{du}{dy} = k \lambda^k y^{k-1} \Rightarrow dy = \frac{1}{k \lambda^k y^{k-1}}$

$y = 0 \Rightarrow u = 0, \; y = \infty \Rightarrow u = \infty$

$$= \cancel{k}\cancel{\lambda^k} \int_0^\infty y^{\cancel{k}} e^{-\cancel{\lambda^k}\frac{u}{\cancel{\lambda^k}}} \frac{1}{\cancel{k}\cancel{\lambda^k}\cancel{y^{k-1}}} du = \frac{1}{\lambda} \int_0^\infty u^{\frac{1}{k}+1-1} e^{-u} du = \frac{1}{\lambda} \Gamma\left(1 + \frac{1}{k}\right)$$

$$\sigma^2 := Var[Y] = \cdots = \frac{1}{\lambda^2}\left(\Gamma\left(1+\frac{2}{k}\right) + \Gamma\left(1+\frac{1}{k}\right)^2\right) < \infty$$

$$\hat{k}^{MLE}, \hat{\lambda}^{MLE} = \text{argmax}\left\{\ell(\lambda, k; \vec{y})\right\}$$ HW: just like the logistic rv, there's no closed-form MLE's

By CLT $\bar{Y} \overset{\cdot}{\sim} N\left(\theta, \frac{\sigma^2}{n}\right) \Rightarrow \hat{\theta} = \bar{y}$ is a reasonable estimate

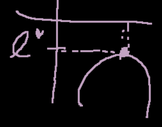
---

Invariance of the MLE thm

Let $\vec{\theta}$ be the parameters for a rv. Let $\vec{\tau} = g(\vec{\theta})$. If $\hat{\vec{\theta}}^{MLE}$ is the MLEs for $\vec{\theta}$ then $g(\hat{\vec{\theta}}^{MLE})$ wil be the MLEs for $\vec{\tau}$.

Proof for if g is a 1:1 function. If not 1:1, it's true but the proof is much more involved. Assume $Y_1, \ldots, Y_n \overset{iid}{\sim}$

$$\sup_{\vec{\tau}}\left\{\ell(\vec{\tau}; \vec{Y})\right\} = \sup_{\vec{\tau}}\left\{\ell(g^{-1}(\vec{\tau}); \vec{Y})\right\}$$
$$= \sup_{\vec{\theta}}\left\{\ell(\vec{\theta}; \vec{Y})\right\}$$



$$\Rightarrow \hat{\theta}^{MLE} = \frac{1}{\hat{\lambda}^{MLE}} \Gamma\left(1 + \frac{1}{\hat{k}^{MLE}}\right)$$

This is a good estimate of "mean survival"

Unfortunately, in most situations, this is impossible to estimate since you need all the y_i's. During your period of observation, you only have a certain amount of time to gather data, call it t_f. All y_i's that realize past this time are... missing (which is called "right-censored" in this context).

Let c be the binary censoring variable. So you observe $\vec{y}, \vec{c}$ as data. If c_i = 1, y_i is missing but known that y_i > t_f. If c_i = 0, then y_i is measured as usual. How do we estimate θ?

$$\mathcal{L}(k, \lambda; \vec{y}, \vec{c}, t_f) = \prod_{\{i : c_i = 0\}} f(y_i; \lambda, k) \prod_{\{i : c_i = 1\}} P(Y > t_f)$$

let $n_0 = \sum \mathbb{1}_{c_i = 0}$,
$+ \; n_1 = \sum \mathbb{1}_{c_i = 1}$
$\overline{\quad n \quad}$

$$= \prod_{\{i : c_i = 0\}} k \lambda^k y_i^{k-1} e^{-\lambda^k y_i^k} \prod_{i=1}^{n_1} e^{-\lambda^k t_f^k}$$

$$\ell(k, \lambda; \vec{y}, \vec{c}, t_f) = n_0 \ln(k) + n_0 k \ln(\lambda) + (k-1) \sum_{\{i : c_i = 0\}} \ln(y_i) - \lambda^k \sum_{\{i : c_i = 0\}} y_i^k - n_1 \lambda^k t_f^k$$

$\ell' \overset{set}{=} 0 \quad$ but.......There is no closed form solution to the root of its deritivate, so you need to use an optimizer.

How do you do inference? The monster MLE thm says:

$$\hat{\theta}^{MLE} \overset{\cdot}{\sim} N\left(\theta, \frac{1}{I_n(\theta)}\right) \xrightarrow{\text{slutsky's}} \hat{\theta}^{MLE} \overset{\cdot}{\sim} N\left(\theta, \frac{1}{I_n(\hat{\theta}^{MLE})}\right)$$

How do we get the Fisher Information? With two-dimensional parameters, the Fisher Information is a 2x2 Fisher Information matrix. And then you use the multivariate delta method. Bottom line: inference is doable.

Is there an alternative to using this censoring data? Yes, it's the EM algorithm.

Step 0: you begin with guesses for the missing y_i's e.g. t_f
Step 1: Compute k-hat-hat-mle, lambda-hat-hat-mle (M-step)
Step 2: Compute better guesses for the missing y_i's (E-step)

E[Y_i | Y_i > t_f] which is the same for all i, censored.

Step 3: Repeate steps 1-2 until convergence i.e. your k, lambda estimates don't change too much between iterations.

Today we covered "parametrice survival modeling" since we assumed the times were realized from a known distribution (in our case, the Weibull). What if we have no idea what distribution it is and we still want inference? We need nonparametric survival modeling tools.